JURNAL TEKNOLOGI TERPADU VOL.13 NO. 2 OKTOBER 2025 ISSN 2338 - 6649

Received: April 2025	Accepted: Juli 2025	Published: Oktober 2025

Prediksi Gangguan Panik Menggunakan *Knowledge Discovery in Database* Dengan Algoritma *Gradient Boosting*

Muammar Ramadhani Maulizidan[1*], Muhammad Lucky Hermanto[2], Onky Ardhillah[3], Muhammad Azyumardi Azra[4], Kevin Agustin Purba[5], Umar Rahman Zidan[6], Ken Ditha Tania[7], Allsela Meiriza[8]

[1,2,3,4,5,6,7,8]Fakultas Ilmu Komputer, Universitas Sriwijaya*

**Email: muammarramadhanimaulizidan@gmail.com*

Abstract

Panic disorder is a mental health condition characterized by sudden panic attacks with intense physical and psychological symptoms. Early and accurate detection is crucial to prevent negative impacts on the patient's quality of life. This study aims to develop a prediction model for panic disorder severity using the Knowledge Discovery in Database (KDD) approach with the Gradient Boosting algorithm. The dataset used consists of 100,000 entries of panic disorder symptoms obtained from the Kaggle platform. The preprocessing steps include data cleaning (no missing values), one-hot encoding for categorical variables (gender, family history), standardization of numerical features (age), and class balancing using SMOTE. The Gradient Boosting model was trained with a 70% training and 30% testing data split, and evaluated using accuracy, precision, recall, F1-score, and confusion matrix metrics. The results show that the model achieved 99.67% accuracy, 99.35% precision, 100% recall, and 99.67% F1-score, with only 186 false positive cases. Feature importance analysis identified lifestyle factors such as sleep quality, stress level, and social support as key contributors in panic disorder diagnosis. This study demonstrates that the integration of KDD, SMOTE, and Gradient Boosting effectively improves the accuracy of panic disorder prediction. The model has the potential to serve as the foundation for a decision support system for early detection. Future work includes leveraging real-time data from wearable devices and testing on a broader population to enhance model generalization.

Keywords: Knowledge discovery in database, gradient boosting, panic disorder

Abstrak

Gangguan panik adalah kondisi kesehatan mental yang ditandai oleh serangan panik mendadak dengan gejala fisik dan psikologis intens. Deteksi dini yang akurat penting untuk mencegah dampak negatif terhadap kualitas hidup penderita. Penelitian ini bertujuan mengembangkan model prediksi gangguan panik menggunakan pendekatan Knowledge Discovery in Database (KDD) dan algoritma Gradient Boosting. Dataset yang digunakan terdiri dari 100.000 entri gejala gangguan panik dari platform Kaggle. Proses pra-pemrosesan mencakup pembersihan data (tanpa nilai hilang), one-hot encoding untuk variabel kategorikal (jenis kelamin, riwayat keluarga), standardisasi fitur numerik (usia), serta penyeimbangan kelas dengan SMOTE. Model Gradient Boosting dilatih dengan pembagian 70% data latih dan 30% data uji, dievaluasi dengan metrik akurasi, presisi, recall, F1-score, dan confusion matrix. Hasil menunjukkan model mencapai akurasi 99,67%, precision 99,35%, recall 100%, dan F1-score 99,67%, dengan hanya 186 kasus false positive. Analisis feature importance mengidentifikasi faktor gaya hidup seperti kualitas tidur, tingkat stres, dan dukungan sosial sebagai kontributor utama dalam diagnosis gangguan panik. Penelitian ini membuktikan bahwa integrasi KDD, SMOTE, dan Gradient Boosting efektif meningkatkan akurasi prediksi gangguan panik. Model ini berpotensi menjadi dasar sistem pendukung keputusan untuk deteksi dini. Disarankan penggunaan data real-time dari perangkat wearable dan uji coba pada populasi lebih luas untuk meningkatkan generalisasi model.

Kata kunci: Knowledge discovery in database, gradient boosting, smote, gangguan panik

1. Pendahuluan

Gangguan panik (*PanicDisorder*) merupakan salah satu gangguan kesehatan mental yang ditandai dengan serangan panik mendadak yang mengakibatkan gejala seperti detak jantung yang cepat, sesak nafas, dan perasaan ketakutan yang intens. Gangguan ini memiliki dampak signifikan pada kualitas hidup penderita dan seringkali mengganggu fungsi sosial serta pekerjaan mereka [1].

Gangguan kecemasan seperti gangguan panik dipengaruhi oleh kombinasi berbagai faktor, baik genetik, biologis, psikologis, maupun lingkungan [2]. Penanganan yang efektif, baik melalui psikoterapi maupun farmakoterapi, sangat penting untuk mengurangi gejala dan mencegah kekambuhan gangguan panik. Dalam hal ini, terapi kognitif perilaku dan penggunaan obat-obatan seperti antidepresan dapat membantu mengelola gejala kecemasan yang berlebihan.

Pentingnya deteksi dini dan pengelolaan gangguan panik telah diakui sebagai faktor kunci dalam mengurangi dampak negatif terhadap kualitas hidup penderita. Deteksi dini gangguan kecemasan, yang meliputi gangguan panik, memungkinkan intervensi lebih cepat yang dapat mencegah perkembangan kondisi menjadi lebih parah [3]. Deteksi yang tepat waktu dan akurat tidak hanya membantu dalam meminimalkan gejala, tetapi juga meningkatkan efektivitas pengelolaan kondisi melalui terapi yang lebih terarah. Oleh karena itu, pengelolaan yang berbasis data dan teknologi seperti ini sangat penting untuk meningkatkan pengelolaan gangguan panik di masyarakat.

Penelitian terdahulu yang melibatkan 126 peserta berusia 18 hingga 60 tahun di Mesir mengungkapkan bahwa 26,2%peserta mengalami gejala panik yang parah, sedangkan 73,8% lainnya mengalami gejala yang sangat parah. Sebagian besar respoded melaporkan tingkat kecemasan sedang (48,4%) hingga berat (38,9%), dan sebanyak 98% melaporkan kualitas hidup yang buruk. Analisis regresi menunjukan bahwa gejala panik memiliki hubungan signifikan dengan usia dan jenis kelamin, sementara tingkat pendidikan berkorelasi dengan kualitas hidup. Temuan ini menekankan pentingnya peningkatan akses terhadap layanan kesehatan mental serta kesadaran masyarakat tentang serangan panik. [4].

Studi lain yang meneliti dampak serangan panik terhadap kualitas hidup pasien di sebuah rumah sakit umum di Turki. Penelitian ini melibatkan 119 pasien gangguan panik, dengan evaluasi menggunakan SF-3 Quality of Life Scale dan formulir informasi sosiodemografi. Hasil penelitian menunjukan bahwa 79% pasien mengalami kejadian negatif sebelum serangan panik pertama mereka, seperti konflik interpersonal atau kehilangan signifikan, yang memicu munculnya gejala panik yang berat. selain itu, 51,3% pasien kehilangan salah satu atau kedua orang tua, yang juga berkontribusi pada frekuensi dan intensitas serangan panik. Faktor-faktor seperti usia, pendidikan, dan durasi pengobatan juga mempengaruhi pengalaman serangan panik: pasien yang lebih tua cenderung memiliki fungsi fisik yang sangat rendah dan nyeri yang lebih tinggi, sementara pendidikan yang lebih tinggi membantu mengurangi dampak nyeri tetapi membuat pasien lebih pasif secara sosial. Temuan ini mempengaruhi aspek psikologis, tetapi juga memperburuk kualitas hidup secara keseluruhan [5].

Meskipun penelitian-penelitian tersebut menunjukan potensi besar dalam prediksi serangan panik, masih terdapat kebutuhan untuk mengintegrasikan berbagai variabel seperti demografi, riwayat medis, dan faktor gaya hidup dalam satu model prediksi yang holistik. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model prediksi tingkat keparahan gangguan panik dengan memanfaatkan teknik KDD yang mengintegrasikan variabel-variabel tersebut. Algoritma *Gradient Boosting* ditetapkan untuk menganalisis data dan mengidentifikasi pola kompleks antar variabel. Hasil dari penelitian diharapkan dapat meningkatkan akurasi dalam mengklasifikasikan tingkat keparahan gangguan panik, sehingga dapat berkontribusi

pada pengembangan sistem pendukung keputusan dalam bidang kesehatan mental.

Prediksi dalam *machine learning* adalah proses memperkirakan nilai atau kejadian di masa depan berdasarkan pola yang ditemukan dalam data historis. Proses ini menggunakan variabel-variabel yang ada dalam dataset untuk memperkirakan nilai dari variabel lain yang belum diketahui. Berbeda dengan tebakan yang bersifat spekulatif, prediksi dalam *machine learning* didasarkan pada model analitis dan algoritmik yang dibangun dari data terstruktur. Pendekatan ini memungkinkan sistem menghasilkan peramalan yang lebih akurat dan objektif, karena didasarkan pada pembelajaran dari data, bukan opini atau intuisi. Secara praktis, prediksi sering kali dikaitkan dengan istilah forecasting, yang keduanya bertujuan untuk memperkirakan masa depan, tetapi *forecasting* lebih berfokus pada pemodelan yang berbasis data dan statistik [6].

2. Metode Penelitian

Penelitian ini menggunakan pendekatan *Knowledge Discovery in Database* (KDD) untuk melakukan analisis data. *Knowledge Discovery in Databases* (KDD) merupakan proses untuk mengekstraksi informasi yang bernilai, mudah dimengerti, dan bersifat baru dari kumpulan data yang besar dan kompleks [7]. Proses ini melibatkan serangkaian aktivitas sistematis yang bertujuan untuk menggali informasi baru yang akurat dan berguna dari kumpulan data yang telah ada [8]. Secara umum, proses KDD mencakup beberapa tahapan utama sebagai berikut:

2.1. Data Selection

Penelitian ini menggunakan dataset yang diperoleh dari *platform Kaggle*, yang dikenal sebagai sumber terpercaya bagi para peneliti dan praktisi data *science* untuk mengakses *dataset* berkualitas. *Dataset* yang digunakan berjumlah 100.000 entri yang berisi informasi terkait gejala *Panic Disorder*. *Dataset* ini dipilih karena memiliki relevansi yang kuat dengan tujuan penelitian dan telah divalidasi oleh komunitas riset di *Kaggle*, yang menjadikannya sumber data yang kredibel. Proses seleksi data dilakukan dengan hati-hati untuk memastikan hanya data yang lengkap dan relevan yang digunakan dalam analisis.

2.2. Preprocessing/Cleaning

Tahap *preprocessing* dilakukan untuk memastikan data siap digunakan dalam analisis lebih lanjut. Pada tahap ini, Preprocessing merupakan langkah penting untuk memastikan bahwa data yang akan dianalisis berada dalam kondisi bersih, konsisten, dan terbebas dari kesalahan [9]. berbagai langkah diterapkan untuk mengidentifikasi dan mengatasi masalah yang ada dalam data, seperti data yang hilang, kesalahan *input*, dan data yang tidak konsisten. Beberapa langkah yang dilakukan adalah pembersihan data, *one-hot encoding*, standarisasi, dan penyeimbangan dengan SMOTE. Tahapan ini memastikan bahwa data yang digunakan dalam analisis lengkap dan konsisten, serta sesuai dengan format yang diperlukan untuk pengolahan selanjutnya.

a. Onehot Encoding

One-Hot Encoding merupakan teknik yang digunakan untuk mengubah data kategorikan menjadi representasi numerik dalam bentuk vektor biner. Setiap kategori direpresentasikan sebagai vektor di mana hanya satu elemen yang bernilai 1, sementara elemen lainnya bernilai 0. teknik ini sangat berguna untuk menghindari masalah *Implicit ordering* [10].

b. SMOTE

Dalam penelitian ini, teknik *Synthetic Minority Over-sampling Technique* atau SMOTE digunakan untuk mengatasi ketidakseimbangan kelas pada dataset. Analisis awal menunjukkan bahwa jumlah data yang terindikasi gangguan panik jauh lebih sedikit dibandingkan dengan data tanpa gangguan. Untuk mencegah bias pembelajaran model terhadap kelas mayoritas, data dari kelas minoritas diperluas secara sintetis menggunakan pendekatan SMOTE. Tahapan ini dilakukan untuk memastikan distribusi kelas

menjadi lebih seimbang dan hasil prediksi lebih akurat [11].

2.3. Transformation

Transformasi data adalah tahapan dalam data mining yang bertujuan mengubah data terpilih agar sesuai dengan kebutuhan analisis. Proses ini sangat bergantung pada struktur basis data, model informasi, dan jenis pola yang ingin ditemukan [12]. Transformasi data merupakan tahap penting yang bertujuan untuk mempersiapkan data agar lebih optimal dalam proses analisis dan pembangunan model. Pada tahap ini, berbagai teknik transformasi diterapkan untuk memastikan seluruh atribut memiliki format yang konsisten dan skala yang seragam, sehingga mendukung kinerja algoritma pembelajaran mesin. Teknik yang digunakan antara lain *one-hot encoding* untuk mengonversi atribut kategorikal menjadi representasi numerik biner, serta standarisasi atribut numerik agar setiap fitur memiliki nilai rata-rata nol dan standar deviasi satu. Melalui proses ini, data menjadi lebih terstruktur, homogen, dan siap digunakan dalam tahapan pemodelan selanjutnya.

2.4. Data Mining

Data mining adalah suatu proses untuk menggali informasi yang bermanfaat dari kumpulan data berskala besar, yang menggabungkan berbagai disiplin ilmu seperti analisis data, penemuan pengetahuan dari basis data, serta pendekatan multidisipliner lainnya seperti statistik, visualisasi data, pengenalan pola, dan manajemen database [13]. Proses Data Mining dilakukan dengan mengekstraksi data dari suatu basis data atau sekumpulan data guna menemukan informasi tersembunyi yang terkandung di dalamnya. Pada tahap ini, proses *data mining* dilakukan dengan menerapkan algoritma *Gradient Boosting*. Algoritma ini dipilih karena kemampuannya dalam membangun model prediksi yang akurat dengan menggunakan pendekatan ansambel, di mana beberapa model lemah digabungkan secara bertahap untuk memperbaiki kesalahan prediksi sebelumnya. *Gradient Boosting* terbukti efektif dalam menangani kompleksitas hubungan antar variabel serta meningkatkan kinerja klasifikasi pada berbagai jenis dataset [14].

2.5. Interpretation/Evaluation

Pada tahap ini akan diperoleh hasil akurasi, presisi, *recall*, dan F1-*score*.Sebelum itu, dilakukan pengujian performa model menggunakan confusion matrix. *Confusion Matrix* adalah alat evaluasi kinerja dalam *machine learning* yang digunakan untuk mengevaluasi hasil klasifikasi.[14]. *Confusion Matrix* digunakan untuk menghitung nilai *precision*, *recall*, dan *accuracy* untuk memvisualisasi seberapa baik model dalam memprediksi kelas positif dan negatif [15]. *Confusion Matrix* memudahkan analisis performa model karena membagi hasil prediksi ke dalam 4 kategori yaitu, *True Positive*, *True Negative*, *False Positive*, dan *False Negative*.

$$f1 = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Pola-pola yang dihasilkan dari proses data mining akan dievaluasi dengan membandingkannya terhadap hipotesis yang telah dirumuskan sebelumnya. Tujuannya adalah untuk mendapatkan kesimpulan yang mendekati hasil yang diharapkan atau sesuai dengan hipotesis guna mendukung tahapan berikutnya.

3. Hasil Penelitian

3.1. Data Selection

Untuk memprediksi gangguan panik, digunakan dataset yang berisi 100.000 data terkait gangguan panik yang diambil dari Kaggle.

Tabel 1. Hasil Analisis Pengujian Lab

No.	Uraian	Keterangan
1	Data 1	Keterangan 1
2	Data 2	Keterangan 2
3	Data 3	Keterangan 3
4	Data 4	Keterangan 4

3.2. Pre-processing/Cleaning

Hasil dari proses pembersihan dan penyesuaian data ditunjukkan pada bagian berikut:

a. Pembersih Data

Sebelum melakukan analisis lebih lanjut, tahap awal yang dilakukan adalah eksplorasi data menggunakan fungsi *train_df.info()*, seperti yang ditampilkan pada Gambar 1.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 17 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   Participant ID          100000 non-null  int64
 1   Age                     100000 non-null  int64
 2   Gender                  100000 non-null  object
 3   Family History          100000 non-null  object
 4   Personal History        100000 non-null  object
 5   Current Stressors       100000 non-null  object
 6   Symptoms                100000 non-null  object
 7   Severity                100000 non-null  object
 8   Impact on Life          100000 non-null  object
 9   Demographics            100000 non-null  object
 10  Medical History         74827 non-null   object
 11  Psychiatric History     75079 non-null   object
 12  Substance Use           66626 non-null   object
 13  Coping Mechanisms       100000 non-null  object
 14  Social Support          100000 non-null  object
 15  Lifestyle Factors       100000 non-null  object
 16  Panic Disorder Diagnosis  100000 non-null  int64
dtypes: int64(3), object(14)
memory usage: 13.0+ MB
```

Gambar 1. Eksplorasi awal *train_df.info()*

Berdasarkan hasil eksplorasi awal seperti pada Gambar 1 bahwa beberapa kolom seperti *Medical History*, *Psychiatric History*, dan *Substance Use* memiliki jumlah data yang terlihat lebih sedikit dari total entri, yaitu 100.000. Jumlah yang lebih sedikit tersebut disebabkan oleh adanya nilai *'None'* yang dituliskan secara eksplisit sebagai teks. Nilai ini digunakan untuk menunjukkan bahwa responden tidak memiliki riwayat pada kategori tersebut. *Missing value* dalam sebuah dataset diberikan tanda seperti NA atau dibiarkan kosong [16].

Dengan demikian, bisa disimpulkan bahwa nilai '*None*' yang muncul bukanlah data kosong, melainkan termasuk kategori yang sah dan bermakna dalam konteks analisis. Oleh karena itu, nilai tersebut tidak dianggap sebagai *missing value*. Secara keseluruhan, data ini tidak memiliki nilai yang benar-benar hilang, sehingga proses data *cleaning* untuk menangani *missing values* tidak diperlukan.

b. Encoding Data Kategorikal

Dilakukan proses *one-hot encoding* untuk atribut kategorikal menggunakan *one-hot encoding*, agar dapat digunakan dalam algoritma *Gradient Boosting*. Pada tabel 2 di tampilkan contoh data yang sudah di*encoding*

Tabel 2. Hasil *one-hot encoding* data kategorikal

Gende_ Female	Gender_ Male	…	Family Hystory_ No	Family History _Yes
0	1	…	0	1
0	1	…	1	0
…	…	…	…	…
0	1	…	1	0
1	0	…	0	1

Tabel 1 menunjukkan hasil dari proses *one-hot encoding* pada data kategorikal, di mana setiap kategori diubah menjadi representasi biner (0 atau 1). Sebagai contoh, atribut *Gender* dan *Family History* dikonversi menjadi beberapa kolom baru yang mewakili setiap kategori. Proses ini penting untuk memungkinkan algoritma seperti *Gradient Boosting* mengolah atribut non-numerik secara efektif. Hal ini dikuatkan dengan adanya penelitian lain yang membandingkan akurasi metode *one-hot encoding* dengan label encoding dalam prediksi linear regresi pada teks label yang dilakukan perubahan menjadi numerik. Kemudian didapatkan hasil R-*Squared* untuk Label *Encoding* 0.54 dan R-*Squared* untuk *One-Hot Encoding* 0.85 [17].

c. Standarisasi Fitur Numerik

Kolom *age* sebagai fitur numerik distandarisasi menggunakan *StandardScaler*. pada tabel x di tampilkan data *age* yang sudah dilakukan standarisasi

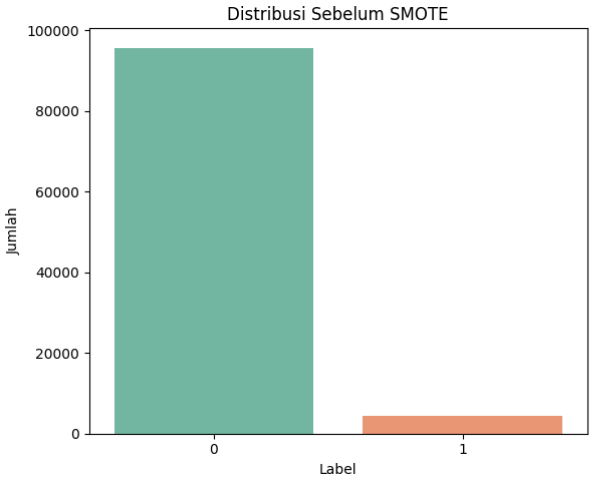
Tabel 3. Standarisasi nilai *age*

Age.	*Age standardization*
38	-0,005393
51	0,162094
32	0,054083
64	-0,017924

Tabel 3 memperlihatkan hasil standarisasi kolom numerik *age* menggunakan metode *StandardScaler*, yang mengubah nilai usia menjadi skor z (nilai yang telah dikurangi rata-rata dan dibagi standar deviasi). Hal ini dilakukan agar skala data menjadi seragam, sehingga algoritma tidak bias terhadap fitur dengan skala lebih besar.

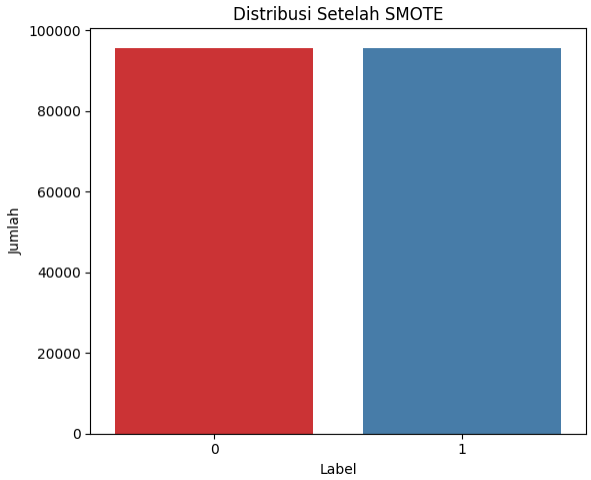
d. Penyeimbangan Kelas

Data awal menunjukkan adanya ketidakseimbangan kelas pada label target, di mana jumlah data penderita gangguan panik jauh lebih sedikit dibandingkan dengan yang tidak. Untuk mengatasi masalah ini, dilakukan oversampling menggunakan metode SMOTE untuk menyeimbangkan distribusi kelas. Gambar 2 menunjukkan distribusi data sebelum penerapan SMOTE.



Gambar 2. Distribusi data sebelum SMOTE

Pada Gambar 3 ditampilkan distribusi data yang sudah di SMOTE. Setelah SMOTE diterapkan, jumlah data pada kedua kelas menjadi seimbang, sehingga model dapat mempelajari representasi dari kedua kelas dengan proporsi yang setara. Kondisi ini, sebagaimana ditampilkan pada Gambar 3, sangat dibutuhkan agar model dapat menghasilkan prediksi yang adil dan akurat untuk seluruh kelas, bukan hanya yang dominan.



Gambar 3. Distribusi data setelah SMOTE

3.3. Transformation

Pada tahap ini, seluruh data yang telah melalui proses *encoding* dan standarisasi digabungkan menjadi satu set fitur yang siap digunakan dalam proses pemodelan. Hasil transformasi tersebut ditunjukkan pada Tabel 4 berikut.

Tabel 4. Hasil *transformation*

Gender_Female	***Gender_Male***	***...***	***Age***	***Panic Disorder Diagnosis***
0	1	...	-0,005393	1
0	1	...	0,162094	0
...	...	...	...	...
0	1	...	0,054083	0
1	0	...	-0,017924	0

Tabel di atas menunjukkan beberapa sampel data setelah melalui tahap pra-pemrosesan. Proses ini mencakup *one-hot encoding* untuk atribut kategorikal dan standarisasi untuk atribut numerik, sehingga seluruh fitur berada dalam format yang seragam dan siap digunakan oleh algoritma pembelajaran mesin.

3.4. Data *Mining*

Setelah tahap transformasi, dilakukan pembangunan model klasifikasi untuk memprediksi gangguan panik. Seluruh nilai fitur dinormalisasi ke dalam rentang 0–1 guna menyamakan skala antar atribut. Model dilatih menggunakan algoritma *Gradient Boosting* dengan pembagian data training sebesar 70% dan data testing sebesar 30%. Penyeimbangan kelas dilakukan terlebih dahulu untuk mengatasi ketimpangan antara peserta yang terindikasi dan tidak terindikasi gangguan panik. Klasifikasi didasarkan pada skor total dari seluruh fitur, dengan ambang batas yang ditentukan secara empiris melalui perhitungan rata-rata atau nilai tengah skor. Kelas 1 menunjukkan individu yang terindikasi mengalami gangguan panik, sedangkan kelas 0 menunjukkan individu yang tidak terindikasi.

3.5. Interpretation/Evaluation

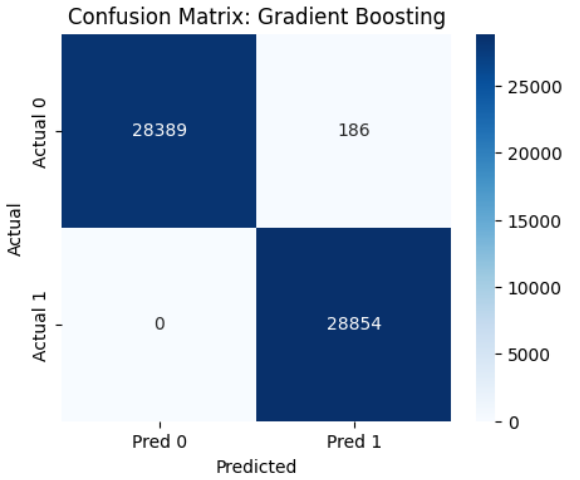
Setelah dilakukan pelatihan dengan data yang telah diseimbangkan, proses evaluasi dilakukan untuk mengukur kinerjanya dalam mengklasifikasikan individu berdasarkan skor total fitur. Evaluasi dilakukan menggunakan data testing sebesar 30%, dengan metrik yang mencakup akurasi, presisi, *recall*, dan F1-*score*. Hasil dari evaluasi tersebut disajikan pada Tabel 4 berikut.

Tabel 4. Evaluasi kinerja model *gradient boosting*

Metrik	Nilai
Akurasi	0.9967
Precision	0,9935
Recall	1.00
F1-Score	0,9967

Hasil evaluasi pada Tabel 4 menunjukkan bahwa model *Gradient Boosting* memiliki performa yang sangat baik, dengan akurasi sebesar 0.9967 yang mencerminkan tingkat prediksi benar yang sangat tinggi secara keseluruhan. Nilai *recall* sempurna (1.00) menunjukkan bahwa semua kasus gangguan panik berhasil terdeteksi dengan benar, sementara nilai F1-*score* sebesar 0.9967 menunjukkan bahwa model memiliki keseimbangan tinggi antara presisi dan recall. Hal ini menandakan bahwa model mampu mengenali kasus dengan sangat akurat sekaligus meminimalkan kesalahan klasifikasi.

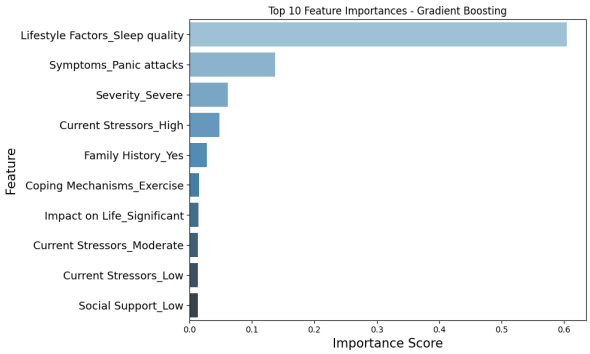
Sebagai pembanding, penelitian lain yang menerapkan label *encoding* dan melakukan tuning hyperparameter dengan hyperopt sebelum menerapkan metode gradient boosting untuk memprediksi keberhasilan telemarketing menghasilkan nilai precision sebesar 94,91% [18]. Menandakan hasil penelitian ini lebih baik dikarenakan mendapatkan nilai precision sebesar 0,9935 dengan recall 1.00.



Gambar 4. Hasil *confusion matrix*

Confusion matrix pada Gambar 4 menggambarkan bahwa model berhasil memprediksi seluruh data kelas positif (penderita gangguan panik) dengan tidak ada kesalahan klasifikasi (*false negative* = 0). Sementara itu, hanya terdapat 186 kasus *false* positive, yaitu kasus non-panik yang salah diklasifikasikan sebagai panik. Angka ini sangat kecil dibandingkan total data, sehingga dapat dikatakan bahwa model sangat presisi dan sensitif terhadap kasus gangguan panik.

Selain mengevaluasi performa model, interpretasi dilakukan untuk memahami fitur-fitur apa saja yang paling berkontribusi terhadap prediksi gangguan panik. Gambar 5 menunjukkan hasil analisis *feature importance* dari model *Gradient Boosting*, yang mengidentifikasi beberapa variabel dengan pengaruh dominan dalam proses prediksi gangguan panik.



Gambar 5. Top 10 faktor gangguan panik

Dari hasil interpretasi ini, dapat disimpulkan bahwa faktor gaya hidup seperti kualitas tidur, tingkat stres, dan dukungan sosial sangat mempengaruhi kondisi gangguan panik. Temuan ini tidak hanya memperkuat literatur sebelumnya dalam psikologi klinis, tetapi juga dapat digunakan sebagai dasar dalam merancang intervensi preventif dan sistem pendukung untuk deteksi dini gangguan panik.

4. Kesimpulan

Penelitian ini berhasil mengembangkan model prediksi gangguan panik menggunakan pendekatan *Knowledge Discovery in Database* (KDD) dengan algoritma *Gradient Boosting* yang dioptimalkan melalui metode SMOTE untuk mengatasi ketidakseimbangan kelas. Model yang dihasilkan menunjukkan performa yang sangat baik dengan akurasi sebesar 99,996%, recall sempurna (1,00), dan F1-*score* sebesar 0,9996, yang mengindikasikan kemampuan tinggi dalam mengklasifikasikan individu yang berisiko mengalami gangguan panik. Selain itu, analisis *feature importance* menunjukkan bahwa faktor gaya hidup seperti kualitas tidur, tingkat stres, dan dukungan sosial memiliki kontribusi signifikan dalam diagnosis gangguan panik.

5. Saran

Untuk meningkatkan kualitas dan relevansi penelitian di masa depan, disarankan agar studi selanjutnya mengeksplorasi pemanfaatan data *real-time* yang diperoleh dari perangkat *wearable*. Penggunaan data semacam ini berpotensi memberikan gambaran kondisi psikologis secara lebih akurat dan terkini. Selain itu, penting pula untuk melakukan pengujian model pada populasi yang lebih luas dan beragam guna meningkatkan generalisasi hasil. Langkah ini diharapkan dapat mendorong penerapan model dalam sistem deteksi dini serta dukungan kesehatan mental berbasis teknologi.

6. Daftar Pustaka

[1] K. Vitoasmara, F. V. Hidayah, N. I. Purnamasari, R. Y. Aprillia, dan L. D. Dewi A., “Gangguan Mental (Mental Disorders),” Student Research Journal, vol. 2, no. 3, pp. 57–68, Jun. 2024,

[2] A. A. G. A. D. A. P. Putra et al., “Generalized Anxiety Disorder (GAD): A Literature Review,” Jurnal Biologi Tropis, vol. 24, no. 1b, pp. 697–703, 2024

[3] T. Anjarsari, I. R. I. Astutik, dan U. Indahyanti, “Deteksi Dini Gangguan Kecemasan Menggunakan Metode Naïve Bayes,” JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), vol. 7, no. 4, pp. 1198–1210, Des. 2022

[4] A. I. Miqdadi, C. M. Chan, M. Alhadidi, T. L. Yoong, and K. O. Hui, “Assessing the Level of Panic Symptoms, Anxiety, and Quality of Life Among People Experiencing Panic Attacks,” Journal of Psychosocial Nursing and Mental Health Services, vol. 0, no. 0, pp. 1–10, Feb. 2025.

[5] N. G. Bayrak and M. Batmaz, "Evaluation of Patients’ Quality Life of Diagnosed with Panic Disorder," EJONS Int. J. on Mathematic, Engineering and Natural Sciences, vol. 8, no. 2, pp. 243–252, 2024.

[6] D. W. Sari dan M. M. Wahyudi, “Analisis dan Perbandingan Algoritma Prediksi dalam Mengetahui Perkiraan Peningkatan Jumlah Kasus COVID-19 di Indonesia dengan Metodologi CRISP-DM,” Jurnal ResearchGate, Jan. 2021.

[7] F. M. A. Sofyan, A. Voutama, Y. Umaidah, “Penerapan Algoritma C4.5 untuk Prediksi Penyakit Paru-paru

Menggunakan RapidMiner," JATI (Jurnal Mahasiswa Teknik Informatika), Apr. 2023.

[8] D. C. P. Buani, "Deteksi Dini Penyakit Diabetes dengan Menggunakan Algoritma Random Forest," Evolusi: Jurnal Sains dan Manajemen, vol. 12, no. 1, pp. 1–8, Mar. 2024.

[9] S. Nachwa, D. K. Harahap, E. T. Pardede, M. N. Ramadhani, S. A. Putri, E. Rositiani, K. D. Tania, A. Meiriza, "Pendekatan Klasifikasi Dalam Knowledge Discovery Untuk Analisis Sentimen Berbasis Aspek Pada Ulasan Bandara Sultan Mahmud Baruddin II Di Google Maps," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 9, no. 3, Jun. 2025.

[10] J. A. Samuels, "One-Hot Encoding and Two-Hot Encoding: An Introduction," Preprint, Jan. 2024.

[11] L. Ma, J. Xu, J. H. D. Cho, E. Korpeoglu, S. Kumar and K. Achan, "NEAT: A Label Noise-resistant Complementary Item Recommender System with Trustworthy Evaluation," 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 469-479, doi: 10.1109/BigData52589.2021.9671870.

[12] M. A. A. Leza, N. W. Utami, P. A. C. Dewi, "Prediksi Prestasi Siswa SMAS Katolik Santo Yoseph Denpasar Berdasrkan Kedisiplinan Dan Tingkat Ekonomi Orang Tua Menggunakan Metode Knowledge Discovery In Database Dan Algoritma Regresi Linier Berganda," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 8, no. 1, Feb. 2024.

[13] I. A. Nikmatun, I. Waspada, "Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," Jurnal SIMETRIS, vol. 10, no. 2, pp. 1–8, Apr. 2021.

[14] W. I. Rahayu, C. Prianto, dan E. A. Novia, "Perbandingan Algoritma K-Means dan Naïve Bayes untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan pada PT. Pertamina (Persero)," Jurnal Teknik Informatika, vol. 13, no. 2, pp. 2252–4983, Apr. 2021.

[15] P. Florek dan A. Zgdański, "Benchmarking State-of-the-Art Gradient Boosting Algorithms for Classification," arXiv preprint, arXiv:2305.17094, May 2023.

[16] A. Zuhairah, Penerapan algoritma Random Forest, Support Vector Machines (SVM) dan Gradient Boosted Tree (GBT) untuk deteksi penipuan (fraud detection) pada transaksi kartu kredit, Bachelor's thesis, Fakultas Sains dan Teknologi, UIN Syarif Hidayatullah Jakarta, 2022.

[17] C. Herdian, A. Kamila, dan I G. A. M. Budidarma, "Studi Kasus Feature Engineering Untuk Data Teks: Perbandingan Label Encoding dan One-Hot Encoding Pada Metode Linear Regresi," Teknologia : Jurnal Ilmiah, vol. 15, no. 1, Jan. 2024.

[18] S. E. Suryana, B. Warsito, and S. Suparti, "Penerapan Gradient Boosting Dengan Hyperopt Untuk Memprediksi Keberhasilan Telemarketing Bank," Jurnal Gaussian, vol. 10, no. 4, pp. 617–623, 2021.